JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 69, 607–620 (1979)

Existence of Optimal Stationary Policies in Deterministic Optimal Control*

DIMITRI P. BERTSEKAS

Department of Electrical Engineering and Computer Science, Mass. Institute of Technology, Cambridge, Mass. 02139

AND

STEVEN E. SHREVE

Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19711

Submitted by M. Aoki

This paper considers deterministic discrete-time optimal control problems over an infinite horizon involving a stationary system and a nonpositive cost per stage. Various results are provided relating to existence of an ϵ-optimal stationary policy, and existence of an optimal stationary policy assuming an optimal policy exists.

1. Problem Formulation and Main Results

The question whether it is possible to restrict attention to stationary policies in deterministic and stochastic optimal control over an infinite horizon has received considerable attention in view of the fact that stationary policies are much easier to implement than nonstationary ones, and can often be computed by means of efficient algorithms. The question is also highly nontrivial and not as yet completely resolved. The purpose of this paper is to provide an analysis of the deterministic optimal control case. The results obtained are to some extent different in nature than those known for the corresponding stochastic case as will be explained in the sequel.

Consider a stationary deterministic system

$$x_{k+1} = f(x_k, u_k), \qquad k = 0, 1, \ldots \tag{1}$$

where x_k and u_k, $k = 0, 1, \ldots$ are elements of given nonempty sets S and C

* This work was supported by the National Science Foundation under Grant NSF ENG 74-20091.

0022-247X/79/060607–14$02.00/0

referred to as the *state space* and *control space* respectively. The function $f: S \times C \to S$ is given. For each $x \in S$ we are given a nonempty set $U(x) \subset C$ referred to as the *control constraint set at* x. Let M be the set of all functions $\mu: S \to C$ such that $\mu(x) \in U(x)$ for all $x \in S$. Let Π be the set of all sequences $(\mu_0, \mu_1, \ldots)$ such that $\mu_k \in M$, $k = 0, 1, \ldots$. An element of Π is referred to as a *policy*. A policy of the form $(\mu, \mu, \ldots)$ where $\mu \in M$ is referred to as a *stationary policy*.

Let $\alpha \in (0, 1]$ be a scalar and $g: S \times C \to [-\infty, 0]$ be a function. We refer to α as the *discount factor* and to g as the *cost per stage*. For each $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ and $x_0 \in S$ define

$$J_\pi(x_0) = \sum_{k=0}^{\infty} \alpha^k g[x_k, \mu_k(x_k)] \tag{2}$$

where x_k, $k = 0, 1, \ldots$ is generated from x_0 and π via the equation

$$x_{k+1} = f[x_k, \mu_k(x_k)], \quad k = 0, 1, \ldots. \tag{3}$$

The function $J_\pi: S \to [-\infty, 0]$ is referred to as the *cost function associated with* π. For a stationary policy $\pi = (\mu, \mu, \ldots)$ we also write J_μ in place of J_π.

Define the *optimal cost function* $J^*: S \to [-\infty, 0]$ by

$$J^*(x) = \inf_{\pi \in \Pi} J_\pi(x), \qquad \forall x \in S.$$

If $J_\pi(x) = J^*(x)$ for a $\pi \in \Pi$ and $x \in S$ we say that π is *optimal at* x. If $J_\pi(x) = J^*(x)$ for all $x \in S$ we say that π is *optimal*. If for an $\epsilon > 0$ and $\pi \in \Pi$ we have for all $x \in S$

$$J_\pi(x) \leqslant J^*(x) + \epsilon \qquad \text{if} \qquad J^*(x) > -\infty$$

$$J_\pi(x) \leqslant -\frac{1}{\epsilon} \qquad \text{if} \qquad J^*(x) = -\infty$$

we say that π is *ϵ-optimal*.

Our results are given in the following two propositions, the proofs of which are provided in the next section.

PROPOSITION 1. *Assume that for each $x \in S$ there exists a policy that is optimal at x.*

(a) *If $\alpha = 1$, then there exists an optimal stationary policy.*

(b) *If $\alpha < 1$ and $J^*(x) > -\infty$ for all $x \in S$, then there exists an optimal stationary policy.*

PROPOSITION 2. *If S is a countable set, and $\alpha = 1$, then for every $\epsilon > 0$ there exists an ϵ-optimal stationary policy.*

The assumption $J^*(x) > -\infty$ for all $x \in S$ cannot be eliminated in Proposition (1b) as the following counterexample shows:

Counterexample 1. Let $\alpha < 1$, $S = \{0\}$, $U(0) = (-\infty, 0]$, $g(0, u) = u$. Then $J^*(0) = -\infty$ and there exists an optimal nonstationary policy — for example $\mu_k(0) = -\alpha^{-k}$, $k = 0, 1,\ldots$. However for every stationary policy $(\mu, \mu,\ldots)$ we have $J_\mu(0) = [1/(1 - \alpha)]\,\mu(0)$.

Also when $\alpha < 1$ the conclusion of Proposition 2 need not hold, even when S is countable. To see this, consider the following counterexample.

Counterexample 2. *Let* $\alpha \in (0, 1)$ and $S = \{-1, 0, 1, 2,\ldots\}$. Consider a control space consisting of two elements s and c ($C = \{s, c\}$). The control s may be viewed as a stopping action that drives the system from any state to state -1 which may be viewed as a termination state. The control c may be viewed as a continuation action which drives the system from any nonnegative state i to state $i + 1$. Thus we have

$$f(x, u) = -1, \qquad \text{if} \quad x = -1 \quad \text{or} \quad u = s$$

$$f(x, u) = x + 1, \qquad \text{otherwise.}$$

If the system is in the termination state or the continuation action c is chosen, then the cost incurred is zero. There is a cost$(\alpha^i - 1)/\alpha^i$, $i = 0, 1,\ldots$ if the stopping action s is chosen at state i. Thus we have

$$g(x, u) = 0, \qquad \text{if} \quad x = -1 \quad \text{or} \quad u = c$$

$$g(i, s) = \frac{\alpha^i - 1}{\alpha^i}, \qquad \forall i = 0, 1,\ldots.$$

Given a stationary policy $\pi = (\mu, \mu,\ldots)$ there are two possibilities: either $\mu(x) = c$ for all $x = 0, 1,\ldots$ in which case $J_\mu(x) = 0$ for all $x \in S$, or else there is a smallest nonnegative integer, say $\bar{\imath}$, such that $\mu(\bar{\imath}) = s$, in which case we have

$$J_\mu(i) = \frac{\alpha^{\bar{\imath}} - 1}{\alpha^i}, \qquad i = 0, 1,\ldots, \bar{\imath}$$

and in particular

$$J_\mu(\bar{\imath}) = 1 - \frac{1}{\alpha^{\bar{\imath}}}.$$

On the other hand it is easy to see that

$$J^*(i) = -\frac{1}{\alpha^i}, \qquad \forall i = 0, 1,\ldots$$

Hence we have

$$J_\mu(\bar{\imath}) = J^*(\bar{\imath}) + 1$$

and the stationary policy $(\mu, \mu, \ldots)$ is not ϵ-optimal if $\epsilon < 1$.

It may be possible to eliminate the countability assumption in Proposition 2. We have neither a proof of this fact nor a counterexample disproving it. It can be seen from our method of proof that if Proposition 2 can be shown for the case where J^* is uniformly bounded below (S not necessarily countable) then the result holds for the general case. For particular classes of problems with uncountable S, such as the stopping problems considered by Dubins and Savage, it is possible to show ([5], p. 60) existence of an ϵ-optimal policy under the assumption that J^* is uniformly bounded below. Our method of proof can thus be used to show that for such problems there exists an ϵ-optimal stationary policy even when J^* is unbounded and/or infinite.

We note that Proposition 1 holds but Proposition 2 fails to hold when $g(x, u) \geqslant 0$ (instead of $g(x, u) \leqslant 0$) for all $(x, u) \in S \times C$ (see [2], Propositions 5.1, 5.4). It can be shown, however, that if $\alpha < 1$ and $g(x, u) \geqslant 0$ for all $(x, u) \in S \times C$ then there exists an ϵ-optimal policy for every $\epsilon > 0$ ([2] Proposition 5.1). Thus the situation regarding existence of ϵ-optimal stationary policies is quite different for the cases $g \geqslant 0$ and $g \leqslant 0$.

The questions considered in this paper have received attention in the works of Dubins and Savage [5], Blackwell [3], [4], and Ornstein [6]. In all these papers the discount factor was taken to be unity ($\alpha = 1$). Furthermore, with the exception of [5], the problems considered in these works are stochastic in nature, i.e., the system evolution contains a stochastic parameter and the expected value of the infinite sum of costs per stage is minimized. The presence of a discount factor less than unity, and of a stochastic element in the problem affect strongly the existence of stationary optimal policies and this constitutes our motivation for restricting attention to deterministic problems while allowing $\alpha < 1$.

More specifically, Proposition (1a) has been given by Ornstein ([6], p. 568) for the case where $J^*(x) > -\infty$ for all $x \in S$, $\alpha = 1$, and a stochastic parameter taking values in a countable space is present in the system equation. Related results have been given by Blackwell [4] for stochastic problems with countable state space and bounded cost per stage, and Dubins and Savage ([5], p. 60) for a special type of gambling problem. Blackwell [4] gives an example showing that for a stochastic problem the assumption $J^*(x) > -\infty$ for all $x \in S$ is essential in order for an optimal stationary policy to exist. Our contribution here consists of showing that this assumption is unnecessary when the problem is deterministic. Furthermore while the proofs of Ornstein and Dubins and Savage are based on Zorn's lemma, our proof is constructive at least for the case where J^* takes finite values everywhere. We have to resort to Zorn's lemma however for the case where $J^*(x) = -\infty$ for some $x \in S$. Our constructive proof of Proposition (1a) can be modified to show part (b) of Proposition 1

which is a new result, and apparently cannot be proved by arguments such as those used by Ornstein and Dubins and Savage.

Proposition 2 has been shown by Ornstein ([6], p. 564) for the case where S is countable, $\alpha = 1$, a stochastic element is present, and J^* is a uniformly bounded function. Ornstein [6] and Blackwell [3] provide examples showing that the countability and boundedness assumptions are both necessary in the presence of a stochastic element. Dubins and Savage ([5], p. 60) show that for a special type of deterministic gambling problem the state space S can be taken to be uncountable but boundedness of J^* is still assumed and used substantively in the proof. Our contribution consists of showing that if S is countable there exists an ϵ-optimal stationary policy even when J^* is unbounded or even infinite. The fact that this need not be true if $\alpha < 1$ is a somewhat surprising and thus far unreported result.

2. Proofs of Propositions 1 and 2

Proof of Proposition 1. Consider the set S_f of states where J^* takes finite values

$$S_f = \{x \in S \mid J^*(x) > -\infty\}. \tag{4}$$

We first assume that $S_f = S$. Subsequently we prove part (a) by extending the proof to the general case where we may have $S_f \neq S$.

A sequence $\{x_0, u_0, x_1, u_1, \ldots\}$ is said to be *admissible* if $u_k \in U(x_k)$ and $x_{k+1} = f(x_k, u_k)$ for all $k = 0, 1, \ldots$. An admissible sequence is said to be *optimal* if $J^*(x_0) = \sum_{k=0}^{\infty} \alpha^k g(x_k, u_k)$. An admissible sequence is said to be *thrifty* if

$$J^*(x_k) = g(x_k, u_k) + \alpha J^*(x_{k+1}), \qquad k = 0, 1, \ldots. \tag{5}$$

From Bellman's equation ([1], Chapter 6, Proposition 8) we have for all $x \in S$

$$J^*(x) = \inf_{u \in U(x)} \{g(x, u) + \alpha J^*[f(x, u)]\} \tag{6}$$

so that an alternative definition of a thrifty sequence can be based on the relation

$$g(x_k, u_k) + \alpha J^*[f(x_k, u_k)] = \inf_{u \in U(x_k)} \{g(x_k, u) + \alpha J^*[f(x_k, u)]\}, \qquad \forall k = 0, 1, \ldots. \tag{7}$$

We have the following lemma (where we assume $S_f = S$):

Lemma 1. *An admissible sequence* $\{x_0, u_0, x_1, u_1, \ldots\}$ *is optimal if and only if it is thrifty and* $\lim_{k \to \infty} \alpha^k J^*(x_k) = 0$.

Proof. If $\{x_0, u_0, x_1, u_1, \ldots\}$ is optimal, then we have for $n = 1, 2, \ldots$

$$J^*(x_0) = \sum_{k=0}^{\infty} \alpha^k g(x_k, u_k) = \sum_{k=0}^{n-1} \alpha^k g(x_k, u_k) + \alpha^n \sum_{k=0}^{\infty} \alpha^k g(x_{n+k}, u_{n+k})$$

and optimality of $\{x_0, u_0, x_1, u_1, \ldots\}$ implies that $\sum_{k=0}^{\infty} \alpha^k g(x_{n+k}, u_{n+k}) = J^*(x_n)$. Hence for $n = 1, 2, \ldots$

$$J^*(x_0) = \sum_{k=0}^{n-1} \alpha^k g(x_k, u_k) + \alpha^n J^*(x_n). \tag{8}$$

It follows that $\lim_{n\to\infty} \alpha^n J^*(x_n) = 0$ and for all $k = 0, 1, \ldots$

$$J^*(x_k) = g(x_k, u_k) + \alpha J^*(x_{k+1}).$$

Hence $\{x_0, u_0, \ldots\}$ is thrifty.

Conversely if $\{x_0, u_0, x_1, u_1, \ldots\}$ is thrifty then (5) holds and implies that (8) also holds. Since $\alpha^n J^*(x_n) \to 0$ it follows that $J^*(x_0) = \sum_{k=0}^{\infty} \alpha^k g(x_k, u_k)$ and $\{x_0, u_0, x_1, u_1, \ldots\}$ is optimal. Q.E.D.

For a scalar β with $0 < \beta < 1$ to be specified further later define

$$A_n = \{x \in S \mid -\beta^n \leqslant J^*(x)\}. \tag{9}$$

We have $\bigcup_{n=-\infty}^{\infty} A_n = S$ and for each $x \in S$ such that $J^*(x) < 0$ there is a unique integer denoted $n^*(x)$ such that

$$x \in A_{n^*(x)} - A_{n^*(x)+1} = \{y \in A_{n^*(x)} \mid y \notin A_{n^*(x)+1}\}. \tag{10}$$

We now prove parts (a), (b) assuming $S_f = S$.

Case 1 ($\alpha = 1$). Let β be any scalar in (0, 1) and define for $x_0 \in A_n - A_{n+1}$ $m^*(x_0) = \min\{m \geqslant 1 \mid$ a thrifty sequence$\{x_0, u_0, \ldots\}$ exists for which $x_m \in A_{n+1}\}$. If $\{x_0, u_0, x_1, u_1, \ldots\}$ is an optimal sequence then $J^*(x_n) \to 0$ so that $m^*(x_0)$ is well defined as a positive integer. Let $\{x_0, u_0, \ldots\}$ be a thrifty sequence for which $x_{m^*(x_0)} \in A_{n+1}$ and define $\mu(x_0) = u_0$. In this way μ is defined on $\bigcup_{n=-\infty}^{\infty} A_n = \{x \in S \mid J^*(x) < 0\}$. If $J^*(x_0) = 0$, let $\{x_0, u_0, \ldots\}$ be any admissible sequence and define $\mu(x_0) = u_0$. We show that the stationary policy $\pi = (\mu, \mu, \ldots)$ is optimal.

For $x_0 \in A_n - A_{n+1}$ let $x_0, u_0, \ldots$ be a thrifty sequence for which $u_0 = \mu(x_0)$ and $x_{m^*(x_0)} \in A_{n+1}$. Either

(a) $m^*(x_0) = 1$, i.e., $x_1 \in A_{n+1}$, $n^*(x_1) \geqslant n + 1$, or

(b) $m^*(x_0) \geqslant 2$.

In case (b) we have $x_1 \in A_n - A_{n+1}$ and in view of the definition of $m^*(\cdot)$ we obtain

$$m^*(x_1) = m^*(x_0) - 1.$$

It follows that if $x_0 \in S$ and $x_0, u_0, \ldots$ is a sequence generated by the stationary policy $\pi = (\mu, \mu, \ldots)$, then $m^*(x_k) = 1$ for infinitely many k's, or else $J^*(x_k) = 0$ for some k (in which case π is optimal at x_0). If $J^*(x_k) \neq 0$ for all k we have $n^*(x_0) = n$, $n^*(x_1) = n + \delta_1$, $n^*(x_2) = n + \delta_1 + \delta_2, \ldots$ where $\delta_k \geqslant 0$, and for infinitely many k's, $\delta_k \geqslant 1$. Hence

$$\lim_{k\to\infty} n^*(x_k) = \infty$$

and since $-\beta^{n^*(x_k)} \leqslant J^*(x_k)$ we obtain $\lim_{k\to\infty} J^*(x_k) = 0$. It follows from Lemma 1 that π is optimal at x_0.

Case 2 ($\alpha < 1$). Let $\beta = \alpha$ and define for $x_0 \in A_n - A_{n+1}$

$$m^*(x_0) = \min\{m \geqslant 1 \mid \text{a thrifty sequence } \{x_0, u_0, x_1, u_1, \ldots\} \text{ exists for which } x_m \in A_{n-m+1}\}.$$

If $\{x_0, u_0, x_1, u_1, \ldots\}$ is an optimal sequence then $\alpha^n J^*(x_n) \to 0$ so that for m sufficiently large we have $-\alpha^{n+1} \leqslant \alpha^m J^*(x_m)$ and $x_m \in A_{n-m+1}$. It follows that $m^*(x_0)$ above is well defined and is a positive integer for each $x_0 \in S$. Let $\{x_0, u_0, \ldots\}$ be a thrifty sequence for which $x_{m^*(x_0)} \in A_{n-m^*(x_0)+1}$ and define $\mu(x_0) = u_0$. In this way μ is defined on $\bigcup_{n=-\infty}^{\infty} A_n = \{x \in S \mid J^*(x) < 0\}$. If $J^*(x_0) = 0$, define $\mu(x_0)$ as in Case 1. We show that the stationary policy $\pi = (\mu, \mu, \ldots)$ is optimal.

For $x_0 \in A_n - A_{n+1}$ let $\{x_0, u_0, x_1, u_1, \ldots\}$ be a thrifty sequence for which $u_0 = \mu(x_0)$ and $x_{m^*(x_0)} \in A_{n-m^*(x_0)+1}$. Either

(a) $m^*(x_0) = 1$, i.e., $x_1 \in A_n$, $n^*(x_1) \geqslant n$, or

(b) $m^*(x_0) \geqslant 2$.

In case (b) we have $J^*(x_1) < -\alpha^n$ and from (5) we obtain also $-\alpha^n \leqslant J^*(x_0) \leqslant \alpha J^*(x_1)$. Hence

$$-\alpha^{n-1} \leqslant J^*(x_1) < -\alpha^n$$

and $x_1 \in A_{n-1} - A_n$, i.e., $n^*(x_1) = n - 1$. Let $\hat{x}_0 = x_1$, $\hat{u}_0 = u_1, \ldots, \hat{x}_{m^*(x_0)-1} = x_{m^*(x_0)}, \ldots$. Since $x_{m^*(x_0)} \in A_{n-m^*(x_0)+1}$ we obtain $\hat{x}_{m^*(x_0)-1} \in A_{(n-1)-[m^*(x_0)-1]+1}$. Hence

$$m^*(x_1) \leqslant m^*(x_0) - 1.$$

It follows from the preceding analysis that if $x_0 \in S$ and $x_0, u_0, \ldots$ is the sequence generated by the stationary policy $\pi = (\mu, \mu, \ldots)$, then $m^*(x_k) = 1$ for infinitely many k's or else $J^*(x_k) = 0$ for some k (in which case π is optimal at x_0). If $J^*(x_k) \neq 0$ for all k we have $n^*(x_0) = n$, $n^*(x_1) = n - 1 + \delta_1$, $n^*(x_2) = n - 2 + \delta_1 + \delta_2, \ldots$ where $\delta_k \geqslant 0$ and for infinitely many k's, $\delta_k \geqslant 1$. Consequently

$$\lim_{k\to\infty}[n^*(x_k) + k] = \infty. \tag{11}$$

Now we have by definition

$$-\alpha^{n^*(x_k)} \leqslant J^*(x_k)$$

and hence

$$\lim_{k\to\infty}[-\alpha^{n^*(x_k)+k}] \leqslant \lim_{k\to\infty} \alpha^k J^*(x_k) \leqslant 0. \tag{12}$$

Combining (11), (12) and the fact $0 < \alpha < 1$ we obtain $\lim_{k\to\infty} \alpha^k J^*(x_k) = 0$. Since $\{x_0, u_0, x_1, u_1, \ldots\}$ is a thrifty sequence it follows from Lemma 1 that it is also optimal and hence π is optimal at x_0.

We now turn to the proof of part (a) for the case where $S_f \neq S$, i.e., when $J^*(x) = -\infty$ for at least one $x \in S$. Given a $\mu \in M$ and a set $\Omega \subset S$ (possibly empty) we say that a policy $\pi = (\mu_0, \mu_1, \ldots)$ is *optimal, μ-stationary and closed on Ω* if, respectively

$$J_\pi(x) = J^*(x), \qquad \mu_k(x) = \mu(x), \qquad \forall x \in \Omega, \quad k = 0, 1, \ldots$$

and furthermore for every $x_0 \in \Omega$ the sequence of states $\{x_0, x_1, \ldots\}$ generated via the system equation using π belongs to Ω, i.e., $x_k \in \Omega$ for all k if $x_0 \in \Omega$ and $x_{k+1} = f[x_k, \mu(x_k)]$ for all k.

We have the following crucial lemma:

LEMMA 2. *Assume that for each $x \in S$, there exists a policy that is optimal at x. If $\pi = (\mu_0, \mu_1, \ldots)$ is optimal, μ-stationary and closed on $\Omega \subset S$, then given any $x_0 \notin \Omega$ there exists a function $\bar{\mu} \in M$, a set $\bar{\Omega} \supset \Omega \cup \{x_0\}$ and a policy $\bar{\pi}$ which is optimal, $\bar{\mu}$-stationary and closed on $\bar{\Omega}$ and $\mu(x) = \bar{\mu}(x)$ for all $x \in \Omega$.*

Proof. We distinguish two cases:

Case 1 ($J^*(x_0) > -\infty$). Define

$$\bar{\Omega} = S_f \cup \Omega$$

where $S_f = \{x \mid J^*(x) > -\infty\}$. Since if the initial state is in S_f all subsequent states under any admissible policy will belong to S_f, by the result proved earlier we can find a stationary policy $\tilde{\mu}$ which is optimal at every point in S_f. Define

$$\bar{\mu}(x) = \tilde{\mu}(x) \qquad \text{if} \qquad x \in S_f - \Omega$$
$$\bar{\mu}(x) = \mu(x) \qquad \text{if} \qquad x \notin S_f - \Omega$$

and consider the stationary policy $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$. Then $\bar{\pi}$ is optimal, $\bar{\mu}$-stationary and closed on $\bar{\Omega}$ and $\mu(x) = \bar{\mu}(x)$ for all $x \in \Omega$.

Case 2 ($J^*(x_0) = -\infty$). Let $\{x_0, u_0, x_1, u_1, \ldots\}$ be an optimal sequence. There are two possibilities:

(a) $x_k \in \Omega$ *for some* $k \geqslant 1$. Let $\bar{k} = \min\{k \geqslant 1 \mid x_k \in \Omega\}$ and consider

the finite sequence $\{x_0, x_1, \ldots, x_{k-1}\}$. If x_0 appears only once in this sequence we consider the sequence $\{x_1, \ldots, x_{k-1}\}$ and apply the process below. If x_0 appears more than once, i.e., the system returns to x_0, let $\bar{n} \leqslant k-1$ be the last integer for which $x_0 = x_{\bar{n}}$. If $\sum_{i=0}^{\bar{n}-1} g(x_i, u_i) < 0$ then a finite subset $\tilde{S} \subset \{x_0, x_1, \ldots, x_{\bar{n}-1}\}$ can be delineated in which no state appears more than once and a function $\tilde{\mu} \in M$ can be obtained such that $\tilde{\mu}(x_i) = u_i$, $J_{\tilde{\mu}}(x_i) = J^*(x_i) = -\infty$ for all $x_i \in \tilde{S}$. Under this policy when we start at x_0 we traverse all states in $\tilde{S}$ one by one and return to x_0 while incurring strictly negative cost in each cycle. The stationary policy $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ for which $\bar{\mu}(x) = \tilde{\mu}(x)$ for $x \notin \Omega$ and $\bar{\mu}(x) = \mu(x)$ for $x \in \Omega$ is optimal, $\bar{\mu}$-stationary and closed on $\bar{\Omega} = \Omega \cup \tilde{S}$ and satisfies the requirement in the lemma. If $\sum_{i=0}^{\bar{n}-1} g(x_i, u_i) = 0$ we replace the sequence $\{x_0, x_1, \ldots, x_{k-1}\}$ by the sequence $\{x_{\bar{n}}, x_{\bar{n}+1}, \ldots, x_{k-1}\}$. Suppose now that $x_{\bar{n}+1}$ occurs more than once in the sequence $\{x_{\bar{n}+1}, \ldots, x_{k-1}\}$. Then as before, we can either construct the desired policy $\bar{\pi}$ or else reduce the sequence $\{x_{\bar{n}+1}, \ldots, x_{k-1}\}$ to one in which $x_{\bar{n}+1}$ occurs only once. We continue doing this, and at the end of the process, i.e., when we reach x_{k-1}, we will either have obtained a policy $\bar{\pi}$ satisfying the requirement of the lemma or else we will have a subsequence $\tilde{S}$ of $\{x_0, x_1, \ldots, x_{k-1}\}$ containing x_0 and x_{k-1} in which each state appears only once. Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ be any policy statisfying $\bar{\mu}(x_i) = u_i$ for each $x_i \in \tilde{S}$ and $\bar{\mu}(x) = \mu(x)$ for all $x \in \Omega$. Then $\bar{\pi}$ is optimal, $\bar{\mu}$-stationary and closed on $\bar{\Omega} = \Omega \cup \tilde{S}$ and $\bar{\mu}(x) = \mu(x)$ for all $x \in \Omega$.

(b) $x_k \notin \Omega$ *for all* k. If a sequence $\{x_0, x_1, \ldots, x_{\bar{n}}\}$ exists for which $x_0 = x_{\bar{n}}$ and $\sum_{i=0}^{\bar{n}-1} g(x_i, u_i) < 0$, then the desired policy $\bar{\pi}$ can be constructed as in (a). Otherwise, the state returns to x_0 infinitely often but $g(x_k, u_k) = 0$ for every k, or else the state returns to x_0 only finitely often. The former case is really not possible, since we have assumed $J^*(x_0) = -\infty$. In the latter case, we let $\bar{n} = \max\{k \geqslant 0 \mid x_k = x_0\}$ and replace $\{x_0, \ldots, x_{\bar{n}}\}$ by $\{x_{\bar{n}}\}$. We apply the same process to $\{x_{\bar{n}+1}, x_{\bar{n}+2}, \ldots\}$ and continue this way so that we obtain either a finite subsequence of $\{x_0, x_1, \ldots\}$ and a stationary policy defined as above which satisfies the requirement of the lemma, or else an infinite subsequence $\tilde{S}$ of $\{x_0, x_1, \ldots\}$ with first element x_0 in which each state appears only once. Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ be any policy satisfying $\bar{\mu}(x_i) = u_i$ for $x_i \in \tilde{S}$ and $\bar{\mu}(x) = \mu(x)$ for all $x \in \Omega$. Then $\bar{\pi}$ is optimal, $\bar{\mu}$-stationary and closed on $\bar{\Omega} = \Omega \cup \tilde{S}$ and $\bar{\mu}(x) = \mu(x)$ for all $x \in \Omega$. Q.E.D.

Let $\mathscr{P}$ be the set of ordered pairs (μ, Ω), where $\mu \in M$, $\Omega \subset S$, and $\pi = (\mu, \mu, \ldots)$ is closed and optimal on Ω. Define an equivalence relation on $\mathscr{P}$ by

$$(\mu, \Omega) \sim (\mu', \Omega') \Leftrightarrow \Omega = \Omega' \qquad \text{and} \qquad \mu(x) = \mu'(x) \qquad \forall x \in \Omega,$$

and partially order the set $\mathscr{R}$ of equivalence classes in $\mathscr{P}$ by the relation

$$[(\mu, \Omega)] \leqslant [(\mu', \Omega')] \Leftrightarrow \Omega \subset \Omega' \qquad \text{and} \qquad \mu(x) = \mu'(x) \qquad \forall x \in \Omega,$$

where $[(\mu, \Omega)]$ denotes the equivalence class containing (μ, Ω). Let $\bar{\mathscr{R}} \subset \mathscr{R}$ be totally ordered and define

$$\bar{\Omega} = \cup \{\Omega \mid [(\mu, \Omega)] \in \bar{\mathscr{R}} \text{ for some } \mu \in M\}.$$

There exists $\bar{\mu} \in M$ such that whenever $[(\mu, \Omega)] \in \bar{\mathscr{R}}$ and $x \in \Omega$, we have $\bar{\mu}(x) = \mu(x)$. The equivalence class $[(\bar{\mu}, \bar{\Omega})]$ is an upper bound for $\bar{\mathscr{R}}$. It follows from Zorn's Lemma that there exists a maximal element in $\mathscr{R}$, i.e., there exists $\mu^* \in M$ and $\Omega^* \subset S$ such that $\pi^* = (\mu^*, \mu^*, \ldots)$ is closed and optimal on Ω^*, and it is not possible to find a $\bar{\mu}^* \in M$ and $\bar{\Omega}^* \subset S$ such that $\pi^* = (\bar{\mu}^*, \bar{\mu}^*, \ldots)$ is closed and optimal on $\bar{\Omega}^*$, Ω^* is properly contained in $\bar{\Omega}^*$ and $\bar{\mu}^*(x) = \mu^*(x)$ for every $x \in \Omega^*$. It follows from Lemma 2 that $\Omega^* = S$, and hence π^* is an optimal stationary policy. This proves part (a). Q.E.D.

Proof of Proposition 2. We will use the following lemma which may be easily deduced from Theorem C of Ornstein [6].

LEMMA 3. *If S is countable, $\alpha = 1$, and there exists a scalar $\beta \in (-\infty, 0]$ such that $\beta \leqslant J^*(x)$, $\forall x \in S$, then for every $\epsilon > 0$ there exists an ϵ-optimal stationary policy.*

Lemma 3 proves Proposition 2 for the case where J^* is uniformly bounded below. It can be used to prove Proposition 2 for the case where $J^*(x) > -\infty$, $\forall x \in S$, as the following lemma shows.

LEMMA 4. *If S is countable, $\alpha = 1$, and $J^*(x) > -\infty$, $\forall x \in S$, then for every $\epsilon > 0$ there exists an ϵ-optimal stationary policy.*

Proof. Define for $n = 0, 1, \ldots$

$$S_n = \left\{ x \in S \;\middle|\; -\frac{(n+1)\epsilon}{2} < J^*(x) \leqslant -\frac{n\epsilon}{2} \right\}.$$

We have $S = \bigcup_{n=0}^{\infty} S_n$. Furthermore if $x_0 \in S_n$ and $\{x_0, u_0, x_1, u_1, \ldots\}$ is any admissible sequence then $x_k \in \bigcup_{i=0}^{n} S_i$ for all $k = 0, 1, \ldots$, i.e., there is no control sequence that can drive the system from x_0 to a state in a set S_k with $k > n$. Thus the optimal cost function of the corresponding deterministic problems where the state space is restricted to be $\bigcup_{i=0}^{n} S_i$, $n = 0, 1, \ldots$ is the function J^* restricted to $\bigcup_{i=0}^{n} S_i$. Select for each $n = 0, 1, \ldots$ a $\mu_n \in M$ satisfying

$$J_{\mu_n}(x) \leqslant J^*(x) + \frac{\epsilon}{2^{n+1}}, \qquad \forall x \in \bigcup_{i=0}^{n} S_i.$$

This is possible by Lemma 3 since J^* is bounded on $\bigcup_{i=0}^{n} S_i$. Define $\mu \in M$ by means of

$$\mu(x) = \mu_n(x) \quad \text{if} \quad x \in S_n.$$

We will show that $(\mu, \mu, \ldots)$ is ϵ-optimal.

Let $x_0 \in S_n$ and let $\{x_k\}$ be the sequence generated by μ, i.e.,

$$x_{k+1} = f[x_k, \mu(x_k)], \qquad k = 0, 1, \ldots.$$

Then

$$\{x_k\} \subset \bigcup_{i=0}^{n} S_i .$$

There are two possibilities:

(a) $x_k \in S_0$, $\forall k \geqslant \bar{k}$ where $\bar{k}$ is some positive integer

(b) $x_k \in S_{\bar{\jmath}}$, $\forall k \geqslant \bar{k}$ where $\bar{k}$, $\bar{\jmath}$ are positive integers with $1 \leqslant \bar{\jmath} \leqslant n$.

We will show that case (b) cannot occur. Indeed if (b) occurred then

$$J_\mu(x_k) = J_{\mu\bar{\jmath}}(x_k), \qquad \forall k \geqslant \bar{k}$$

while for all $k \geqslant \bar{k}$

$$J_\mu(x_k) = J_{\mu\bar{\jmath}}(x_k) \leqslant J^*(x_k) + \frac{\epsilon}{2^{\bar{\jmath}+1}} \leqslant -\frac{\bar{\jmath}\epsilon}{2} + \frac{\epsilon}{2^{\bar{\jmath}+1}} \leqslant -\frac{\epsilon}{4} .$$

On the other hand we clearly have $J_\mu(x_k) \to 0$, which leads to a contradiction.

Hence case (a) occurs and there exists a $\bar{k}$ such that

$$\mu(x_k) = \mu_0(x_k), \qquad \forall k \geqslant \bar{k}.$$

Now let $N_0, N_1, \ldots, N_m$ and $n_0, n_1, \ldots, n_{m-1}$ be positive integers such that $N_0 < N_1 < \cdots < N_m$, $n_{m-1} < n_{m-2} < \cdots < n_0 < n$ and

$$\begin{aligned}
&x_k \in S_n , && \forall 0 \leqslant k \leqslant N_0 \\
&x_k \in S_{n_0} , && \forall N_0 + 1 \leqslant k \leqslant N_1 \\
&\quad\vdots && \quad\vdots \\
&x_k \in S_{n_{m-1}} , && \forall N_{m-1} + 1 \leqslant k \leqslant N_m \\
&x_k \in S_0 , && \forall N_m + 1 \leqslant k.
\end{aligned}$$

We have

$$J_\mu(x_0) = \sum_{i=0}^{N_0} g[x_i, \mu_n(x_i)] + \sum_{i=N_0+1}^{N_1} g[x_i, \mu_{n_0}(x_i)] + \cdots$$

$$+ \sum_{i=N_{m-1}+1}^{N_m} g[x_i, \mu_{n_{m-1}}(x_i)] + J_{\mu_0}(x_{N_m+1}).$$

We also have

$$\sum_{i=0}^{N_0} g[x_i, \mu_n(i)] + J_{\mu_n}(x_{N_0+1}) \leqslant J^*(x_0) + \frac{\epsilon}{2^{n+1}}$$

$$\sum_{i=N_0+1}^{N_1} g[x_i, \mu_{n_0}(x_i)] + J_{\mu_{n_0}}(x_{N_1+1}) \leqslant J^*(x_{N_0+1}) + \frac{\epsilon}{2^{n_0+1}}$$

$$\cdots \qquad \cdots$$

$$\sum_{i=N_{m-1}+1}^{N_m} g[x_i, \mu_{n_{m-1}}(x_i)] + J_{\mu_{n_{m-1}}}(x_{N_m+1}) \leqslant J^*(x_{N_{m-1}+1}) + \frac{\epsilon}{2^{n_{m-1}+1}}$$

$$J_{\mu_0}(x_{N_m+1}) \leqslant J^*(x_{N_m+1}) + \frac{\epsilon}{2}.$$

It follows from the relations above that

$$J_\mu(x_0) \leqslant J^*(x_0) + \epsilon. \qquad \text{Q.E.D.}$$

We shall also need the following lemma:

LEMMA 5. *If* $\alpha = 1$ *and* $J^*(x) = -\infty$ *for all* $x \in S$, *then there exists an optimal stationary policy.*

Proof. Given any $x_0 \in S$ there exists an admissible sequence $\{x_0, u_0, x_1, u_1, \ldots\}$ and a nonnegative integer N_{x_0} such that $\sum_{k=0}^{N_{x_0}} g(x_k, u_k) < -1$. It follows that for any $x_0 \in S$ there exists a policy optimal at x_0. Hence, by Proposition 1(a), there exists an optimal stationary policy. Q.E.D.

We are ready now to prove Proposition 2. Consider the sets

$$S_f = \{x_0 \in S \mid J^*(x_0) > -\infty\}$$

$$S_\infty = \{x_0 \in S \mid J^*(x_0) = -\infty\}$$

$$S_\epsilon = \Big\{x_0 \in S_\infty \mid \text{there exists } u \in U(x_0), \text{ such that } f(x_0, u) \in S_f, \text{ and}$$
$$g(x_0, u) + J^*[f(x_0, u)] \leqslant -\frac{1}{\epsilon} - \epsilon\Big\}$$

$$\hat{S}_\epsilon = \{x_0 \in S_\infty \mid \text{there exists an admissible sequence} \{x_0, u_0, \ldots\} \text{ such that}$$
$$x_{\bar{k}} \in S_\epsilon \text{ for some } \bar{k} \geqslant 0\}$$

$$\tilde{S}_\epsilon = \{x_0 \in S_\infty \mid x_0 \notin \hat{S}_\epsilon\}.$$

For $x_0 \in \tilde{S}_\epsilon$ consider the set

$$\tilde{U}(x_0) = \{u \in U(x_0) \mid f(x_0, u) \in \tilde{S}_\epsilon\}.$$

Clearly $\tilde{U}(x_0)$ is nonempty for all $x_0 \in \tilde{S}_\epsilon$. We claim that

$$\inf\left\{\sum_{k=0}^{\infty} g(x_k, u_k) \mid x_{k+1} = f(x_k, u_k), u_k \in \tilde{U}(x_k), k = 0, 1,...\right\} = -\infty, \quad \forall x_0 \in \tilde{S}_\epsilon. \tag{13}$$

Assume the contrary, i.e., that there exists an $\bar{\epsilon} > 0$ and an $x_0 \in \tilde{S}_\epsilon$ such that for all admissible sequences $\{x_0, u_0, x_1, u_1, ...\}$ for which $x_k \in \tilde{S}_\epsilon$, $u_k \in \tilde{U}(x_k)$, $\forall k = 0, 1,...$ we have

$$\sum_{k=0}^{\infty} g(x_k, u_k) > -\frac{1}{\bar{\epsilon}}.$$

Consider an admissible sequence $\{\bar{x}_0, \bar{u}_0, \bar{x}_1, \bar{u}_1, ...\}$ with $\bar{x}_0 = x_0$ such that

$$\sum_{k=0}^{\infty} g(x_k, u_k) < -\frac{1}{\epsilon} - \epsilon - \frac{1}{\bar{\epsilon}}. \tag{14}$$

Then there must exist a state $\bar{x}_N$ such that $\bar{x}_N \in \tilde{S}_\epsilon$ and $\bar{x}_k \notin \tilde{S}_\epsilon$ for all $k > N$. We have from (14)

$$\sum_{k=0}^{N-1} g(\bar{x}_k, \bar{u}_k) + \sum_{k=N}^{\infty} g(\bar{x}_k, \bar{u}_k) < -\frac{1}{\epsilon} - \epsilon - \frac{1}{\bar{\epsilon}}.$$

Since $\sum_{k=0}^{N-1} g(\bar{x}_k, \bar{u}_k) + 1/\bar{\epsilon} > 0$ it follows that

$$\sum_{k=N}^{\infty} g(\bar{x}_k, \bar{u}_k) < -\frac{1}{\epsilon} - \epsilon.$$

Hence

$$g(\bar{x}_N, \bar{u}_N) + J^*(\bar{x}_{N+1}) < -\frac{1}{\epsilon} - \epsilon,$$

which implies that $\bar{x}_N \in \hat{S}_\epsilon$. Since $\bar{x}_N \in \tilde{S}_\epsilon$ we obtain a contradiction.

Equation (13) together with Lemma 5 implies that there exists a $\tilde{\mu} \in M$ such that

$$J_{\tilde{\mu}}(x) = J^*(x) = -\infty, \qquad \forall x \in \tilde{S}_\epsilon,$$

and $f[x, \tilde{\mu}(x)] \in \tilde{S}_\epsilon$ if $x \in \tilde{S}_\epsilon$.

By Lemma 4 there exists a $\mu_f \in M$ such that

$$J_{\mu_f}(x) \leqslant J^*(x) + \epsilon, \qquad \forall x \in S_f,$$

and $f[x, \mu_f(x)] \in S_f$ if $x \in S_f$.

Consider the sequence of sets $\{\hat{S}_{\epsilon,k}\}$ defined by

$$\hat{S}_{\epsilon,0} = S_\epsilon$$

$$\hat{S}_{\epsilon,k+1} = \{x \in \hat{S}_\epsilon \mid \text{there exists } u \in U(x) \text{ such that } f(x, u) \in \hat{S}_{\epsilon,k}\}.$$

Clearly $\bigcup_{k=0}^{\infty} \hat{S}_{\epsilon,k} = \hat{S}_\epsilon$. Consider a $\hat{\mu} \in M$ having the property

$$f[x, \hat{\mu}(x)] \in S_f, \quad g[x, \hat{\mu}(x)] + J^*[f(x, \hat{\mu}(x))] \leqslant -\frac{1}{\epsilon} - \epsilon \qquad \text{if} \quad x \in S_\epsilon$$

$$f[x, \hat{\mu}(x)] \in \hat{S}_{\epsilon,k} \quad \text{if} \quad x \in \hat{S}_{\epsilon,k+1} \qquad \text{and} \qquad x \notin \bigcup_{j=0}^{k} \hat{S}_{\epsilon,k}.$$

Define $\mu \in M$ by means of

$$\mu(x) = \mu_f(x) \qquad \text{if} \qquad x \in S_f$$

$$\mu(x) = \hat{\mu}(x) \qquad \text{if} \qquad x \in \hat{S}_\epsilon$$

$$\mu(x) = \tilde{\mu}(x) \qquad \text{if} \qquad x \in \tilde{S}_\epsilon.$$

Then it is easy to see that $(\mu, \mu, \ldots)$ is ϵ-optimal. Q.E.D.

References

1. D. P. Bertsekas, "Dynamic Programming and Stochastic Control," Academic Press, New York, 1976.
2. D. P. Bertsekas and S. E. Shreve, "Stochastic Optimal Control: The Discrete Time Case," Academic Press, New York, 1978.
3. D. Blackwell, Positive dynamic programming, *in* "Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability," Vol. 1, pp. 415–418, 1967.
4. D. Blackwell, On stationary policies, *J. Royal Statist. Soc. Ser. A.* **133** (1970), 33–37.
5. L. E. Dubins and L. J. Savage, "How to Gamble if You Must," McGraw–Hill, New York, 1965.
6. D. Ornstein, On the existence of stationary optimal policies, *Proc. Amer. Math. Soc.* **20** (1969), 563–569.